

SYSTEM AND METHOD FOR CALCULATING AN OPTIMUM DISPLAY SIZE FOR A VISUAL OBJECT

5 PRIORITY AND RELATED APPLICATIONS

The present application claims priority to provisional patent application entitled, "VIDEO PROCESSING METHOD WITH GENERAL AND SPECIFIC APPLICATIONS," filed on July 11, 2000 and assigned U.S. Application Serial No. 60/217,301. The present application is also related to non-provisional application
10 entitled, "ADAPTIVE EDGE DETECTION AND ENHANCEMENT FOR IMAGE PROCESSING," (attorney docket number 07816-105003) filed on July 11, 2001 and assigned U.S. Application Serial No. _____; and non-provisional application entitled, "VIDEO COMPRESSION USING ADAPTIVE SELECTION OF GROUPS OF FRAMES, ADAPTIVE BIT ALLOCATION, AND ADAPTIVE REPLENISHMENT,"
15 (attorney docket number 07816-105004) filed on July 11, 2001 and assigned U.S. Application Serial No. _____.

TECHNICAL FIELD

The present invention relates generally to video processing techniques. More particularly, the present invention relates to a method and system for calculating an optimum display size for a visual object.
20

BACKGROUND OF INVENTION

Computer networks such as the internet, local area networks, and wide area
25 networks have become the preferred communication medium for sending and receiving both traditional and new forms of multi-media content. For example, receiving digital video over the internet that is displayed on a computer screen has become very popular. Digital video involves storing, moving, and calculating extremely large volumes of data compared to other kinds of computer files. While the data rate and file size of
30 uncompressed digital video can overwhelm many personal computers and hard disks, compressed video can lower the data rate of the digital video into a range that a particular computer system can handle.

Compression is often defined as the coding of data to save storage space or transmission time. Although data can be coded in digital form for computer processing, data can be coded more efficiently with compression, using fewer bits. Video compression involves the compression of sequences of images. Video compression algorithms typically use the fact that there are usually only small changes from one "frame" to the next so they only need to encode the starting frame and a sequence of differences between frames.

The ISO Committee that generates standards for digital video compression and audio is the Moving Picture Experts Group ("MPEG"). MPEG also refers to the name of this Committee's algorithms for video compression. MPEG-1 was the first MPEG format for compressed video that was optimized for CD-ROM. MPEG-1 was designed for transmission rates of about 1.1 megabits per second achievable with video-CD and CD-i. Another compression algorithm is MPEG-2, which is a variant of the MPEG video and audio compression algorithm in file format that was optimized for broadcast quality video. MPEG-2 was designed to transmit images using progressive coding at 4 megabits per second or higher for use in broadcast digital TV and DVD.

A proposed variant of the MPEG video and audio compression algorithm in file format was MPEG-3. MPEG-3 was intended as an extension of MPEG-2 to cater for HDTV but was eventually merged into MPEG-2. MPEG-4 is another video compression algorithm that was planned for the year of 1998. MPEG-4 extends the earlier MPEG-1 and MPEG-2 algorithms with synthesis of speech and video, fractal compression, computer visualization and artificial intelligence-based image processing techniques.

Many conventional media players that are designed to receive multimedia from the internet, such as the WINDOWS® Media Player and APPLE® Quick Time, utilize the MPEG-1 format for compressed video. The MPEG-1 format utilizes discrete cosine transform ("DCT") and Huffman Coding to remove spatially redundant data within a frame and block-based motion compensated prediction ("MCP") to remove data which is temporally redundant between frames. Audio with the MPEG-1 format is typically compressed using subband encoding. The MPEG-1

algorithm allows better than VHS quality video and almost CD quality audio to be compressed onto and streamed off of a single speed (1x) CD-ROM drive.

While the aforementioned compression techniques may permit the transmission of visual objects over a computer network, such compression techniques can introduce artifacts or distortion in the video as a direct result of the compression techniques. For example, MPEG-1 encoding can introduce artifacts such as blockiness, mosquito noise, color bleeding, and shimmering effects within the video in addition to a lack of detail in quantization effects in audio. While the effects of the compression techniques on the video segments of a multimedia presentation may be negligible in some instances, typically the artifacts and distortion generated by the compression techniques become quite apparent to a user when a visual object such as video is enlarged on a display device.

For example, in many conventional systems utilizing compressed video, the video decoder sets a nominal fixed display size and the user of the video decoder may have options for manually enlarging the display size by activating pixel repetition or interpolation routines in the media player. When a decoded or uncompressed video signal is enlarged, both the raw components of the original uncompressed video signal as well as the distortion caused by the compression techniques are enlarged. In order words, an enlarged image with a conventional media player typically enlarges or amplifies both the original video signal and the distortion.

Consequently, in conventional real time streaming of the video, frequently the size of the video is fairly small relative to the size of the display device. In other words, conventional streamed video is typically not a full screen experience since it is recognized that both the original uncompressed video signal as well as the distortion caused by the compression techniques can be noticeable or recognizable by a user when a large screen rendition of a video object is selected. Therefore, some conventional media players do not even permit the selection for enlarging a video segment. The video segments are typically fixed at small display sizes in order compensate for the artifacts or distortion created by the video compression techniques.

In addition to compression related artifacts discussed above, there also exists packet loss related artifacts that are a result of the transmission speed of a multimedia segment. For example, low bit rates can also contribute to the artifacts or distortion when compressed video is decoded. Low bit rates can be defined as data transmission rates that are at approximately twenty-five percent (25%) of a standard speed. For example, if MPEG-1 is a reference standard, then transmission rates of about 1.1 megabits per second would be the norm for the standard rate. A low rate would be about twenty-five percent (25%) of the normal rate and, therefore, 250 megabits per second could be categorized as a low bit rate.

Referring now Figure 1A, this figure illustrates one exemplary frame 100A of a visual object 105A at a first size relative to the frame's original size. Figure 1A could be referred to as the optimum display size for the exemplary frame 100A since artifacts or distortion are not readily perceived by the user.

Opposite to Figure 1A, Figure 1B illustrates an exemplary frame 100B of the visual object 105B illustrated in Figure 1A at a second size of the frame's original size that is much larger than the first size illustrated in Figure 1A. In Figure 1B, artifacts or distortion such as mosquito noise 110, 115 and the shimmering effects 120 can be perceived in this enlargement of the exemplary frame of a visual object 100B.

While the enlargement techniques of the conventional art may magnify or amplify the artifacts or distortion caused by compression or low transmission rates, enlarging of visual objects as well as reducing the size of visual objects are features that are desirable to the user. In other words, permitting enlargement or reduction in size of a visual object enables efficient use of the work space on a display device. However, as noted above, a balance should be struck between displaying a visual object large enough to be comfortably viewed by a user while at the same time, displaying the visual object at a size that does not permit image degradations due to compression or transmission to be noticeable or easily perceived by the user.

Accordingly, there is a need in the art for a method and system for calculating an optimum display size for a visual object. That is, there is a need in the art for a method and system for automatically selecting the optimal size of a visual object that can be displayed on a display device. Further, there is an need in the art for a method

and system that can determine the best size to display a visual object received from a computer network to a computer screen such that image degradations due to compression or transmission defects or both are less noticeable or not readily perceivable by a user. There is a further need in the art for a method and system that
5 can display a message indicating the optimal display size and receive input as to whether the optimal display size is selected by a user.

SUMMARY OF THE INVENTION

The present invention is generally drawn to a method and system for
10 automatically selecting the optimal size of a visual object that can be displayed on a display device. For example, the present invention can determine the best or optimum size to display a visual object received from a computer network to a computer screen such that image degradations due to compression or transmission defects are less noticeable or perceivable. Typically, the best or optimum size to display a visual
15 object can comprise a size that is also the maximum in which the visual object can be displayed without any degradations or defects.

According to one exemplary aspect, the present invention can automatically perform the sizing of the received visual object. According to another exemplary aspect, the invention can display a message indicating the optimal display size and
20 receive input as to whether the optimal display size is selected by a user. The visual object can comprise one of video, a graphical image, and other like visual objects. The visual object can be transmitted across various types of communication mediums in various formats. The communication mediums can comprise computer networks, wireless networks, infra-red transmissions, electromagnetic couplings, and other
25 similar communication mediums. The formats of the visual objects can include downloadable files, internet streaming, and visual objects retrievable from storage mediums such as CD-ROMs or DVDs. The display devices for the present invention can include a printer, cathode ray tubes (CRTs), liquid crystal displays (LCDs), light emitting diode (LED) displays, and projectors. These display devices can be part of
30 various equipment such as computers (laptops, desktops, handhelds - such as PDAs, etc.), televisions, cellular telephones, and other similar equipment.

According to one aspect, a coding difficulty value can be calculated after compressing a visual object in a conventional format, such as Moving Picture Experts Group (MPEG) format, for computer network transmission. One exemplary coding difficulty value can comprise a harmonic average of the Peak Signal to Noise Ratio (PSNR) for a number of frames taken from the visual object. This coding difficulty value can reflect the presence of visual object segments that will likely be coded poorly. An encoder can calculate the harmonic average of the PSNR based upon the original visual object to be encoded. More specifically, the visual object encoder can calculate the PSNR for each frame of the original visual object or by reviewing a set of frames or a sample of frames and comparing the frame or set of frames to the compressed video. Using the PSNR values for the number of frames evaluated, the encoder can then calculate the harmonic mean of the PSNR values.

The coding difficulty value of the present invention is not limited to the harmonic average of the PSNR. Other coding difficulty values are not beyond the scope of the present invention. For example, the coding difficulty value could comprise the straight average of the PSNR values for a number of frames of visual object or any other subjectively relevant metric.

According to one exemplary aspect of the present invention, once the coding difficulty value is determined, the optimum display size for the encoded visual object can be determined based upon the coding difficulty value and one or more other parameters. More specifically, according to one aspect, the optimum display size can be determined based upon the coding difficulty value and the coding bit rate. With at least these two parameters, the coding bit rate and the coding difficulty value, the optimum display size that can provide the most clear and largest presentation of the encoded visual object can be determined. The optimum display size along with the encoded visual object can then be transmitted to a receiver or decoder side of the system.

The present invention is not limited to the coding difficulty and coding bit rate parameters for determining an optimum display size. Other parameters include, but are not limited to, screen quality such as screen size and screen resolution (dots per inch - dpi) capability.

According to a further exemplary aspect of the present invention, once the coding difficulty value is determined, the coding difficulty value can then be transmitted with the encoded visual object over a communications medium such as a computer network. Next, during reception of the encoded visual object and coding
5 difficulty value, the optimum display size for the encoded visual object can be determined based upon the coding difficulty value and one or more other parameters. More specifically, according to one aspect, the optimum display size can be determined based upon the coding difficulty value and the coding bit rate. With these two parameters, the coding bit rate and the coding difficulty value, the optimum
10 display size that can provide the most clear and largest presentation of the encoded visual object can be determined.

According to one aspect, the optimum display size can be determined by mapping the coding difficulty value and the visual object transmission rate on a graph, where the coding difficulty value can define the x-coordinate and the visual object
15 transmission rate can define the y-coordinate in a Cartesian coordinate system. The Cartesian coordinate system can further comprise a plurality of diagonal zones as well as stair-step type functions. Depending upon where a particular set of coordinates will fall within a particular stair-step type function can provide information as to the amount a particular visual object can be sized. In other words, each stair step type
20 function may be associated with a fixed value defining a size of the visual object relative to a computer screen. For example, each stair step function can comprise a twenty to twenty-five percent increase in size relative to each other and vice-versa. The diagonal zones and stair step functions can be calculated empirically.

Once the optimum display size is calculated for a particular visual object that
25 is to be decoded, the present invention can automatically size the visual object according to this calculation or the calculated display size can be displayed as an option that can be selected. If the option is selected, then the visual object can be decoded and displayed at the calculated optimum display size.

According to another aspect of the present invention, a coding difficulty value
30 can be calculated after a visual object has been compressed in a conventional format, such as Moving Picture Experts Group (MPEG) format, for computer network

transmission. A decoder can review the step size history of the compression format's quantizers. That is, in many compression formats, quantization methods can be used to round off or quantize discrete cosine transform (DCT) coefficients. A step size can comprise the quantization of DCT coefficients. By evaluating the steps sizes for
5 frames of a particular encoded visual object, then the PSNR can be estimated. More specifically, the harmonic average of the PSNR can be estimated or approximated by looking at a mean value of step sizes for a set of frames for an encoded visual object. This estimation of the harmonic average of the PSNR can be derived empirically. Once the estimated harmonic average of the PSNR is determined, the optimum
10 display size can be calculated, as discussed above, by mapping the harmonic average of the PSNR and the visual object transmission rate and determining the zone in which the coordinates are located.

According to a further aspect of the present invention, the optimum display size can be calculated as a function of the coding difficulty value alone.
15 Alternatively, the optimum display size can be calculated as a function of the visual object transmission rate or encoding rate alone.

BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1A illustrates one exemplary frame of a visual object at a first size of
20 its original size according to one exemplary embodiment of the present invention.

Figure 1B illustrates a conventional exemplary frame of the visual object depicted in Figure 1A at a second size of its original size that is larger than the first size.

Figure 1C illustrates an encoding system for calculating a coding difficulty
25 value and for encoding a visual object according to one exemplary embodiment of the present invention.

Figure 1D illustrates a decoding system for calculating a coding difficulty value and for decoding an encoded visual object after receiving the encoded visual object according to another exemplary embodiment of the present invention.

30 Figure 2A illustrates a method for calculating an optimum display size at an encoder, transmitting the calculated optimum display size to a receiver, and rendering

a visual object according to the calculated optimum display size in one exemplary embodiment of the present invention.

Figure 2B illustrates a method for calculating a coding difficulty value at an encoder, transmitting the calculated coding difficulty value to a receiver, and rendering a visual object according to the calculated optimum display size in one exemplary embodiment of the present invention.

Figure 2C illustrates a method for calculating an optimum display size at an encoder based on a coding difficulty value alone, transmitting the calculated optimum display size to a receiver, and rendering a visual object according to the calculated optimum display size in one exemplary embodiment of the present invention.

Figure 2D illustrates a method for calculating a coding difficulty value at an encoder, transmitting the calculated coding difficulty value to a receiver, and rendering a visual object according to the calculated optimum display size based on the coding difficulty value alone in one exemplary embodiment of the present invention.

Figure 3A illustrates a subroutine of Figures 2A, 2B, and 4A for determining the optimum display size of a visual object according to one exemplary embodiment of the present invention.

Figure 3B is a graph illustrating relationships between harmonic peak signal to noise ratios and a visual object transmission rate that may comprise stair step type functions or diagonal zones in accordance with an exemplary embodiment of the present invention.

Figure 4A illustrates a method for calculating an optimum display size at a decoder and rendering visual object according to the calculated optimum display size according to one exemplary embodiment of the present invention.

Figure 4B illustrates a method for calculating an optimum display size at a decoder based upon the visual object transmission rate alone and rendering visual object according to the calculated optimum display size according to one exemplary embodiment of the present invention.

Figure 4C illustrates a method for calculating an optimum display size at a decoder and rendering visual object according to the calculated optimum display size

that is based on a coding difficulty value alone according to one exemplary embodiment of the present invention.

DETAILED DESCRIPTION OF EXEMPLARY EMBODIMENTS

5 The present invention can determine the maximum size to display a visual object received from a computer network to a computer screen such that image degradations due to compression or transmission defects are the least noticeable. The visual object can comprise one of video, a graphical image, and other like visual objects. In one exemplary embodiment, a coding difficulty value can be
10 calculated during compression of a visual object in a conventional format, such as Moving Picture Experts Group (MPEG) format, for computer network transmission. According to another aspect of the present invention, a coding difficulty value can be calculated from a received coded or compressed visual object that was transmitted over a computer network. An optimum display size can be calculated by mapping the
15 harmonic average of the PSNR and the visual object transmission rate and determining the zone in which the coordinates are located.

Illustrative Operating Environment

20 The present invention can be embodied in software that runs on a computer. The processes and operations performed by a computer include the manipulation of signals by a processor and the maintenance of these signals within data structures resident in one or more memory storage devices. For the purposes of this discussion, a process is generally conceived to be a sequence of computer-executed steps leading to a desired result. These steps usually require physical manipulations of physical
25 quantities. Usually, though not necessarily, these quantities take the form of electrical, magnetic, or optical signals capable of being stored, transferred, combined, compared, or otherwise manipulated. It is convention for those skilled in the art to refer to representations of these signals as bits, bytes, words, information, elements, symbols, characters, numbers, points, data, entries, objects, images, files, or the like.
30 It should be kept in mind, however, that these and similar terms are associated with appropriate physical quantities for computer operations, and that these terms are

merely conventional labels applied to physical quantities that exist within and during operation of the computer.

It should also be understood that manipulations within the computer are often referred to in terms such as creating, adding, calculating, comparing, moving, receiving, determining, identifying, populating, loading, executing, etc. that are often associated with manual operations performed by a human operator. The operations described herein can be machine operations performed in conjunction with various input provided by a human operator or user that interacts with the computer.

In addition, it should be understood that the programs, processes, methods, etc. described herein are not related or limited to any particular computer or apparatus. Rather, various types of general purpose machines may be used with the functions described herein. Similarly, it may prove advantageous to construct a specialized apparatus to perform the method steps described herein by way of dedicated computer systems in a specific network architecture with hard-wired logic or programs stored in nonvolatile memory, such as read-only memory.

Referring now to the drawings, in which like numerals represent like elements throughout the several Figures, aspects of the present invention and the illustrative operating environment will be described. Figure 1A illustrates one exemplary frame 100A of a visual object 105A at a first size relative to its original size. The size of the exemplary frame 100A illustrated in Figure 1A can be the optimum display size for the visual object 105A. The optimum size of a visual object 105A can be defined as the size of a visual object 105A at which artifacts or distortion are not literally perceivable or apparent to a user. Stated differently, the optimum display size can be defined as the maximum size at which a visual object 105A can be displayed on the display device without apparent or readily perceivable artifacts or distortion caused by compression or transmission rate or both.

A visual object can comprise one of video, a single graphical image, or other like visual objects. The graphical image can comprise one of a banner advertisement, a photograph, and a graphical object. The video can comprise one of a live television signal and stored video. More specifically, the formats of the visual objects can include downloadable files, internet streaming, and visual objects retrievable from

storage mediums such as CD-ROMs or DVDs. The display devices that can render frame 100A of Figure 1A can comprise printers, monitors for desktop or laptop computers such as cathode ray tubes (CRT) or liquid crystal displays (LCDs), handheld computers such as Personal Digital Assistants (PDAs), cellular telephones, watch displays, and other like devices. Frame 100A can be transmitted over various communication mediums.

For example, frame 100A can be transmitted over a computer network such as a local area network (LAN) or the Internet. However, the present invention is not limited to these communication mediums. Other communication mediums include, but are not limited to, wireless communication links such as radio frequency communications, infra red communications, electromagnetic coupling, and other similar communication mediums.

System for Calculating Optimum Display Size of a Visual Object

Referring now to Figure 1C, this figure illustrates a functional block diagram of an encoding system 100 for calculating a coding difficulty value of a visual object according to one exemplary embodiment of the present invention. The exemplary encoding system 125 can comprise a scene change detector 130, a temporal bit allocator 135, a rate controller 140, and a preprocessing filter 145. The scene change detector 130 is connected to the temporal bit allocator 135. The temporal bit allocator is connected to the rate controller 140. The encoding system 100 receives a video input signal and an audio input signal. The video input is typically a series of digitized images that are linked together in series. The audio input is simply the audio signal that is associated with the series of images making up the video input.

The video input is first passed through a pre-processing filter 145 that, among other things, filters noise from the video input to prepare the input video stream for encoding. The input video stream is then passed to the visual object or video encoder 150. The video encoder 150 compresses the video signal by eliminating irrelevant and/or redundant data from the input video signal. The video encoder 150 may reduce the input video signal to a predetermined size to match the transmission requirements of the encoding system 100. Alternatively, the video encoder 150 may

simply be configured to minimize the size of the encoded video signal. This configuration might be used, for example, to maximize the storage capacity of a storage medium (e.g., hard drive). The video encoder 150 can perform the compression process by reducing the input data rate to a level that can be supported by the storage or transmission medium. It can also translate the compressed bit stream into a signal suitable for either storage or transmission. The video encoder 150 can also calculate the mean square error (MSE) for the one or more frames of the visual object of interest. The video encoder 150 can further calculate a coding difficulty value as will be discussed in further detail below with respect to Figure 2.

In a similar fashion, the audio input is compressed by the audio encoder 160. The encoded audio signal is then passed with the encoded video signal to the video stream multiplexer 165. The video stream multiplexer 165 combines the encoded audio signal and the encoded video signal so that the signals can be separated and played-back substantially simultaneously. After the encoded video and encoded audio signals have been combined, the encoding system outputs the combined signal as an encoded video stream. The encoded video stream is thus prepared for transmission, storage, or other processing as needed by a particular application. Often, the encoded video stream will be transmitted to a decoding system that will decode the encoding video stream and prepare it for subsequent display.

In an exemplary embodiment of the present invention, the video input stream can be further processed prior to encoding. In addition to the pre-processing performed by the pre-processing filter 145, the encoding system 100 of exemplary embodiment of the present invention also can prepare the input video stream for encoding by modifying the input video stream to facilitate compression. The exemplary encoding system 100 includes a picture grouping module or scene change detector 130, a bit allocation module or temporal allocator 135, and a bit rate controller 140.

The picture grouping module 130 can process a video input stream by selecting and classifying I-pictures in the video stream. The picture grouping module 130 can also select and classify P-pictures in the video stream. The picture grouping module 130 can significantly improve the quality of the encoded video stream.

Conventional encoding systems arbitrarily select I-pictures, by adhering to fixed-size picture groups. The coding system 100 is more fully described in co-pending non-provisional patent application entitled, "VIDEO COMPRESSION USING ADAPTIVE SELECTION OF GROUPS OF FRAMES, ADAPTIVE BIT ALLOCATION, AND ADAPTIVE REPLENISHMENT," (attorney docket number 07816-105004) filed on July 11, 2001 and assigned U.S. Application Serial No. _____, the contents of which are hereby incorporated by reference.

The bit allocation module 135 can be used to enhance the quality of the encoded video bit stream by adaptively allocating bits among the groups of pictures defined by the picture grouping module 130 and by allocating bits among the pictures within a given group of pictures. Whereas conventional decoding systems often allocate bits in an arbitrary manner, the allocation module 135 can reallocate bits from the picture groups requiring less video data to picture groups requiring more video data. Consequently, the quality of the encoded video bit stream is enhanced by improving the quality of the groups of pictures requiring more video data for high quality representation.

The bit rate controller 140 can use an improved method of conditional replenishment to further reduce the presence of noise in an encoded video bit stream. Conditional replenishment is a well-known aspect of video data compression. In conventional encoding systems, a picture element or a picture block will be encoded in a particular picture if the picture element or block has changed when compared to a previous picture. Where the picture element or block has not changed, the encoder will typically set a flag or send an instruction to the decoder to simply replenish the picture element or block with the corresponding picture element or block from the previous picture.

The bit rate controller 140 can focus on macroblocks and may condition the replenishment of a macroblock on the change of one or more picture elements and/or blocks within the macroblock, as discussed in co-pending non-provisional patent application entitled, "VIDEO COMPRESSION USING ADAPTIVE SELECTION OF GROUPS OF FRAMES, ADAPTIVE BIT ALLOCATION, AND ADAPTIVE REPLENISHMENT," (attorney docket number 07816-105004) filed on July 11, 2001 and assigned U.S. Application Serial No. _____, the contents of which are hereby incorporated by reference. Alternatively, the bit rate

controller 140 may condition the replacement of a macroblock on a quantification of the change within the macroblock (e.g., the average change of each block) meeting a certain threshold requirement. In any event, the objective of the bit rate controller 140 is to further reduce the presence of noise in video data and to simplify the encoding of a video stream.

The exemplary encoding system 125 may further comprise a display size selector 155. The display size selector 155 can then determine the optimum display size by utilizing the coding difficulty value calculated by the visual object encoder 150 in addition to other parameters as will be discussed in further detail below with respect to Figure 2. The display size selector 155 has been illustrated with dashed lines to indicate that in other exemplary embodiments, this element can be eliminated when the coding difficulty value is transmitted over a communications medium so that a decoder 167 with a display size selector 175 can determine the optimum display size as will be discussed below with respect to the exemplary embodiment illustrated in Figure 2B.

Referring now to Figure 1D, this figure illustrates a functional block diagram of a decoding system 167 for calculating a coding difficulty value after receiving the encoded visual object according to another exemplary embodiment of the present invention. The decoding system 167 receives an encoded video stream as input to a video stream demultiplexer 170. The video stream demultiplexer 170 separates the encoded video signal and the encoded audio signal from the encoded video stream. The encoded video signal is passed from the video stream demultiplexer 170 to the video decoder 172. Similarly, the encoded audio signal is passed from the video stream demultiplexer 170 to the audio decoder 182. The video decoder 172 and a audio decoder 182 expand the video signal and the audio signal to a size that is substantially identical to the size of the video input and audio input.

Those skilled in the art will appreciate that various well-known algorithms and processes exist for decoding an encoded video and/or audio signal. It will also be appreciated that most encoding and decoding processes are lossy, in that some of the data in the original input signal is lost. Accordingly, the video decoder 172 will reconstruct the video signal with some signal degradation is often receivable as flaws

in the output image. The video or visual object decoder 172 can review the step size history of the compressed visual object as will be discussed in further detail below with respect to Figure 4A.

5 The post-processing filter 177 is used to counteract some of the well-known noise found in a decoded video signal. That has been encoded and/or decoded using a lossy process. Examples of such well-known noise types include mosquito noise, salt-and-pepper noise, and blockiness. The conventional post-processing filter 177 includes well-known algorithms to detect and counteract these and other known noise problems. The post-processing filter 177 generates a filtered, decoded video output.
10 The post processing filter 177 can filter the decoded visual object to improve a user's perception of the accuracy of the rendered visual object.

15 However, in one exemplary embodiment, the post-processing filter 177 typically has no impact on the process for determining the optimal display size of the visual object. In other words, the post-processing filter 177, in some exemplary embodiments, does not directly affect the display size algorithm, but helps increase the impact of a decoded visual object. Lowering noise level with a post-processing filter 177 can create a conservative situation for displaying a visual object at a larger size.

20 The display size selector 175 can determine the optimum display size of a visual object by utilizing an estimate of the coding difficulty value based upon the step sizes calculated by the visual object encoder 150 as will be discussed in further detail below with respect to Figure 4A. The visual object renderer 180 can change the size of a visual object by repeating pixels or performing linear extrapolation by using linear filtering methods. The visual object renderer 180 will size the visual
25 object based upon information received from the display size selector 175.

The video output and the audio output (via audio output device 185) may be fed to appropriate ports on a display device, such as a television, or may be provided to some other display means such as a software-based media playback component on a computer. Alternatively, the video output and the audio output may be stored for
30 subsequent display.

As described above, the video decoder 172 decompresses or expands the encoded video signal. While there are various well-known methods for encoding and decoding a video signal, in all of the methods, the decoder must be able to interpret the encoded signal. The typical decoder is able to interpret the encoded signal received from an encoder, as long as the encoded signal conforms to an accepted video signal encoding standard, such as the well-known MPEG1 and MPEG2 standards. In addition to raw video data, the encoder typically encodes instructions to the decoder as to how the raw video data should be interpreted and represented (i.e., displayed).

For example, an encoded video stream may include instructions that a subsequent video picture is identical to a previous picture in a video stream. In this case, the encoded video stream can be further compressed, because the encoder need not send any raw video data for the subsequent video picture. When the decoder receives the instruction, the decoder will simply represent the subsequent picture using the same raw video data provided for the previous picture. Those skilled in the art will appreciate that such instructions can be provided in a variety of ways, including setting a flag or bit within a data stream.

Exemplary Methods for Calculating an Optimum Display Size of a Visual Object

Certain steps in the processes described below must naturally precede others for the present invention to function as described. However, the present invention is not limited to the order of the steps described if such order or sequence does not alter the functionality of the present invention. That is, it is recognized that some steps may be performed before or after other steps without departing from the scope and spirit of the present invention.

Referring now to Figure 2A, this figure illustrates a method for calculating an optimum display size at an encoder, such as encoding system 125, and rendering a visual object according to the calculated optimum display size. Step 205A is the first step of the process for calculating an optimum display size at an encoder. In step 205A, a visual object such as a video segment is encoded according to a

predetermined format. For example, the visual object encoder 150 can encode the received visual object according to the MPEG-1 video compression format.

Next, in step 210A, for each frame or a set of frames or a sample of frames from the original visual object, a coding difficulty value can be calculated as a function of the signal-to-noise ratio (SNR). In step 210A, the peak signal-to-noise ratio (PSNR) of a particular frame can be calculated with the encoder 150 by comparing the original visual object with the encoded visual object. PSNR can be calculated as follows:

$$PSNR = 10 \text{ Log}_{10} \frac{255^2}{\text{noise signal energy}}$$

10 The noise signal energy is defined as the energy measured for a hypothetical signal that is the difference between the encoder input signal and the decoder output signal. Once the peak signal-to-noise ratio is calculated for the frames of interest for the visual object, then the coding difficulty value can be calculated.

There can be two parameters that control the coding difficulty of a display size: (1) the degree of compression which is signified by the data rate; (2) the inherent difficulty of a scene. The difficulty or complexity of a scene can have many different types of measurements associated with it.

The coding difficulty value can comprise a harmonic mean of the peak signal-to-noise ratio (PSNR) for one or more frames taken from the visual object. The harmonic mean of the Peak Signal to Noise Ratio (PSNR) can be used to identify the difficulty of scenes within a visual object. This exemplary coding difficulty value can reflect a presence of visual object segments that will likely be coded poorly during the compression process. The coding difficulty and the signal noise ratio are usually inversely related. In other words, the ease of coding is proportional to the harmonic mean and the coding difficulty is the inverse effect so when the harmonic average is higher, larger display sizes are usually suggested. The harmonic mean of the PSNR values (HM-PSNR) for all the frames in the input sequence can be derived from the following equation:

$$HM - PSNR = 10 \log_{10} \frac{N}{\sum_{i=1}^N \frac{MSE_i}{255^2}}$$

where HM-PSNR = the harmonic average of the peak signal-to-noise ratio; where N = the number of frames of a visual object; and where MSE = the mean square error for each of the frames of the visual object. Basically, step 210 can be performed by the visual object encoder 150.

The harmonic mean is a statistic which can emphasize the weakest link in a chain or series of data points. In the context of the present invention, if different segments of the visual object have different levels of fidelity which is measured by the signal to noise ratio, then the harmonic mean physically emphasizes the lower end. In other words, the harmonic mean of the peak to signal noise ratio usually gives a conservative measure of the weakest link or the video segment which may likely have the poorest quality. If the peak values of the signal to noise ratio (PSNR) are used themselves as the measurement of scene complexity, such values may give an overly optimistic view of how successful the video compression has been. If the arithmetic average of the peak to noise ratio is utilized, such a measurement may give a somewhat optimistic view of how good the overall quality of the compression of the video was performed. Therefore, the harmonic mean is typically referred to as an overall statistic that is conservative in nature that is well suited for determining optimal display sizes for visual objects.

The coding difficulty value of the present invention is not limited to the harmonic mean of the peak signal-to-noise ratio. Other coding difficulty values are not beyond the scope of the present invention. For example, the coding difficulty value could comprise the straight average of the peak signal-to-noise ratio values for a number of frames of a visual object.

Once the coding difficulty value is calculated in step 210A, in routine 215, the optimum display size for the visual object 105A can be determined based upon the coding difficulty value and the transmission rate of the visual object. Further details of routine 215 will be discussed below with respect to Figure 3. Alternatively, as will be discussed below with respect to the exemplary embodiments illustrated in Figures

2C, 2D, 4B, and 4C, the optimum display size can be calculated based upon either the coding difficulty value or the visual object transmission rate. Further, the optimum display size calculation can also evaluate other parameters in addition to the coding difficulty value and visual object transmission rate. For example, the optimum display size can consider the quality of the display device such as its resolution capabilities (dots per inch - dpi) and the size of the display device, just to name a few alternative and exemplary parameters. Those skilled in the art will recognize that other parameters affecting the calculation of the optimum display size are not beyond the scope of the present invention. Calculation of the optimum display size can be biased or weighted appropriately depending upon these other parameters such as display device quality and size.

Next, in step 220A, the encoded visual object and optimum display size value are forwarded over a communications medium. For example, the encoded visual object and optimum display size can be transmitted across a computer network such as the Internet or a local area network. The present invention is not limited to communication mediums comprising computer networks. The communication mediums can include, but are not limited to, wireless networks, infra red transmissions, electromagnetic couplings, and other similar communication mediums.

In step 225A, the received visual object is decoded. Next, in decision step 230A, it is determined whether the visual object is displayed with the calculated optimum display size. Basically, decision step 230A can comprise various options for the system designer. That is, decision step 230A can comprise a prompt for input from a user. In other words, the user could select the optimum display size function as the default size for any visual object received from a network.

Alternatively, decision step 230A can be embodied in options that are dependent upon payment for a particular service. Stated differently, decision step 230A can be connected to the type of service selected and paid for by a user. In one exemplary embodiment, decision step 230A can be conditioned upon the receipt of a form of payment in order to determine the optimum display size for the visual object. Decision step 230A may further comprise a step of displaying the calculated optimum

display size as message so that a user may see the actual calculated optimum display size value.

If the inquiry to decision step 230A is positive, then the "yes" branch is followed to step 235A, in which the visual object with the calculated optimum display size is displayed. If the inquiry to decision step 230A is negative, then the "no" branch is followed to step 240A, in which the visual object is displayed with a default or selected display size activated by the user varied.

Referring now to Figure 2B, this figure illustrates a method for calculating a coding difficulty value at an encoder, such as encoding system 125, transmitting the coding difficulty value over a communications medium, and rendering a visual object according to the calculated optimum display size. As noted above, certain steps can be performed in a different order with out departing from the spirit and scope of the present invention. The process depicted in Figure 2B differs slightly from the process described in Figure 2A.

Only the differences between Figures 2A and 2B will be discussed with respect to Figure 2B. Accordingly, in Figure 2B, step 215A for calculating the optimum display size has been moved to the decoder side or receiver side of the process, as step 227. Step 220A has been modified as step 221B such that the coding difficulty value instead of the calculated optimum display size is transmitted over the communications medium.

Referring now to Figure 2C, this figure illustrates a method for calculating a coding difficulty value and an optimum display size at an encoder, such as encoding system 125, the optimum display size being calculated based on only the coding difficulty value, transmitting the optimum display size over a communications medium, and rendering a visual object according to the calculated optimum display size. As noted above, certain steps can be performed in a different order with out departing from the spirit and scope of the present invention. The process depicted in Figure 2C differs slightly from the process described in Figure 2A.

Only the differences between Figures 2A and 2C will be discussed with respect to Figure 2C. Accordingly, in Figure 2C step 215 has been modified as step 216 to calculate the optimum display size based upon only the coding difficulty value

instead of the combination of the coding difficulty value and the visual object transmission rate as discussed above.

Referring now to Figure 2D, this figure illustrates a method for calculating a coding difficulty value at an encoder, such as encoding system 125, transmitting the coding difficulty value over a communications medium, calculating an optimum display size based on the coding difficulty value alone, and rendering a visual object according to the calculated optimum display size. As noted above, certain steps can be performed in a different order with out departing from the spirit and scope of the present invention. The process depicted in Figure 2D differs slightly from the process described in Figure 2B.

Only the differences between Figures 2B and 2D will be discussed with respect to Figure 2D. Accordingly, in Figure 2D step 227 has been modified as step 228 to calculate the optimum display size based upon only the coding difficulty value instead of both the coding difficulty value and the visual object transmission rate.

Referring now to Figure 3A, this figure illustrates the subroutine 215 of Figure 2A, in which the optimum display size for a visual object is calculated by the display size selector 155. Step 305 is the first step of routine 215 in which the coding difficulty value calculated in step 210 is ascertained. In step 310, the visual object transmission rate is also ascertained. In one exemplary embodiment the visual object transmission rate comprises one or more values measured in units of information per unit of time period. More specifically, the visual object transmission rate can comprise a speed at which binary digits are transmitted. For example, the visual object transmission rate in one exemplary embodiment can comprise bits per second.

In step 315, the coding difficulty value and visual object transmission rate value are mapped to a chart, table, graph, grid or other similar data association mechanism. In one exemplary embodiment, the coding difficulty value defines an X coordinate while the visual object transmission rate defines a Y coordinate in a Cartesian coordinate system. In step 320, from the coordinates identified in step 315, a zone value corresponding to the coding difficulty value and transmission rate can be identified. In step 325, this zone value can be associated with an optimum display

size value such as the percentage in which a visual object should be enlarged. Further details of steps 315, 320, and 325 will be discussed below with respect to Figure 3B.

Referring now to Figure 3B, this figure is a graph illustrating relationships between the harmonic mean for the peak signal-to-noise ratios (HM-PSNR) and the visual object transmission rate (in kilobits per second or kbps). In Figure 3B, the harmonic average of the peak signal-to-noise ratio defines the X-axis while the transmission rate defines the Y-axis. Figure 3B illustrates empirically derived relationships between data transmission rates and a harmonic average of the peak signal-to-noise ratio of a compression technique. The zones A, B, C, D and E can be associated with percent increases of picture size for displaying visual objects. Zone A comprises the region between the X-axis and a first stair step type function 335. Zone B defines a region between the first stair step function 335 and a second stair step function 340.

Zone C defines a region between the second stair step type function 340 and a third stair step type function 345. Similarly, Zone D defines a region between the third stair step type function and a fourth stair step type function 350. And lastly, Zone E defines a region extending above the fourth stair step type function 350.

Each zone of Zones A to E can be associated with a respective percentage increase of the visual object size for each respective zone depicted in Figure 3B. For example, each zone could define a 20% increase in size relative to a neighboring zone. However, the present invention is not limited to the 20% increments. Other increments, higher or lower, are not beyond the scope of the present invention. In Figure 3B, it can be assumed that there are five different display size levels. The original size of an encoded image can be assumed to be one quarter (usually, 320x240) of full size (usually, 640x480). Therefore, Zone A can be interpreted to mean there is no blow-up. That is, the decoder can just display the image at the original display size. The maximum display size in Figure 3B can be 200% of the original size, and this size is associated with zone E. The blow-up factor or percent increase associated with each zone can be as follows: (1-dimensional factors; that is, the actual increased display size will be the square of the following factors): zone A = 1x, zone B = 1.2x, zone C = 1.4x, zone D = 1.7x, and zone E = 2x.

Diagonal Zones I, II, III, IV, V and VI can rationalize the empirical procedure taken to define the graph illustrated in Figure 3B. The diagonal Zones I, II, III, IV, V and VI demonstrate that for a particular sample object, such as Sample Object 1 which is represented by a triangle in Figure 3B, each of the data points for the Sample Object 1 appear to fall within a diagonal zone of II.

The data points for each sample visual object were calculated by taking a visual object such as a video segment and adjusting the transmission rate. Since the transmission rate affects the coding or compression of a visual object, different harmonic means of the PSNRs were derived for the same sample visual object. For example, for a first data point 360 of Sample Visual Object 1, this data point 360 was encoded for a bit rate of approximately 110 kilobits per second. Such a transmission speed yielded a harmonic average of the PSNR of approximately 30. This data point 360 is located in Zone B. Zone B has an associated percent increase value of 20%. Therefore, the decoded visual object is recommended to be displayed at 120% of its original size to provide the optimal size in which artifacts or distortion caused by the transmission rate or compression algorithm or both are less noticeable.

Similarly, for another second data point 365 of Sample Visual Object 1, the sample was encoded for a bit rate of approximately 240 kilobits per second. Such a transmission speed yielded a harmonic average of the PSNR of approximately 32.5. This second data point 365 is located in Zone E. Zone E has an associated percent increase value of 100%. Therefore, the decoded visual object is recommended to be displayed at 200% of its original size to provide the optimal size in which artifacts or distortion caused by the transmission rate or compression algorithm or both are less noticeable.

While only six sample visual objects were used to empirically define the relationships illustrated in the graph of Figure 3B, more or fewer sample objects may be employed without departing from the scope and spirit of the present invention. Further, the present invention is not limited to formatting data into a graph or chart format. That is, the present invention can also take the data illustrated in Figure 3B and place it into a table where the Cartesian coordinates of Figure 3B are associated with respective values of the stair step type functions in a tabular or table format.

Further, when the present invention is embodied in a chart or table format, the process can be automated without departing from the scope and spirit of the present invention.

Referring now to Figure 4A, this figure illustrates an exemplary method for calculating an optimum display size at a decoder and rendering a visual object according to the calculated optimum display size. Step 405A is the first step of the optimum display size calculation process 400A in which an encoded visual object is received from a communications medium such as from across a computer network such as the internet or a local area network. Next, in Step 410A, from each frame, a set of frames, or a sample of frames, or a combination thereof, an average step size is determined from the visual object based on a first transform coefficient. In other words, the decoder 172 can review the step size history of a visual object compression format's quantizers. That is, in many compression formats, quantization methods can be used to round off or quantize Discrete Cosine Transform (DCT) coefficients. A step size can comprise the quantization of DCT coefficients.

Discrete Cosine Transform is a technique for expressing a waveform as a weighted sum of cosines. The method can be expressed as follows: Given data $A(i)$, where i is an integer in the range 0 to $N-1$, the forward DCT (which would be used e.g. by an encoder) is:

$$B(k) = \sum_{i=0}^{N-1} A(i) C_k \cos \left[\frac{(2i+1)k\pi}{2N} \right]$$

where $C_0 = \sqrt{1/N}$ and $C_k = \sqrt{2/N}$ ($k = 1, \dots, N-1$)

The inverse DCT (which would be used e.g. by a decoder) is:

$$AA(i) = \sum_{k=0}^{N-1} B(k) C_k \cos \left[\frac{(2i+1)k\pi}{2N} \right]$$

The main difference between DCT and a discrete Fourier transform (DFT) is that the DFT traditionally assumes that the data $A(i)$ is periodically continued with a period of N , whereas the DCT assumes that the data is continued with its mirror image, then periodically continued with a period of $2N$. Mathematically, this transform pair is exact, i.e. $AA(i) == A(i)$, resulting in lossless coding; only when some of the coefficients are approximated does compression occur.

In Step 415A, similar to Step 410A, for each frame, a set of frames, a sample of frames, or a combination thereof, an average step size based on a second transformed coefficient can be determined from the visual object. In Step 420A, the appropriate function can be calculated with the first and second average step size sets.

5 In one exemplary embodiment, a mean value could be calculated from the first and second average step size sets.

In Step 425A, the final step size parameter can be calculated from the appropriate functions defined above. Once the final step size parameter is calculated in Step 425A, then in Step 430A, a coding difficulty value can be ascertained based upon the final step size parameter. In Step 430A, an estimated harmonic mean of the peak signal to noise ratio can be calculated. Similar to Figure 3B, the relationship between the final step size parameter calculated in Step 425A and the estimated harmonic mean of the peaked signal noise ratio can be empirically determined.

10

Next, in Routine 435A, the optimum display size for the visual object is determined based upon the coding difficulty value (the estimated harmonic mean of the PSNR) and the transmission rate. Subsequently, in Step 440A, the received and coded visual object is decoded. In decision Step 445A, it is determined whether the visual object the calculated optimum display size should be displayed. If the inquiry to decision Step 445A is positive, then the "yes" branch is followed to Step 450A in which the visual object with the calculated optimum display size is displayed. If the inquiry to decision Step 445A is negative, then the "no" branch is followed to Step 455A in which the visual object is displayed with a default or selected display size.

15

20

As noted above with respect to decision Step 230A, decision Step 445A can comprise various options. That is, decision Step 445A can be embodied in a default setting for rendering a visual object. Alternatively, Step 445A can be embodied in pull down menus that enable a user to select whether automated display size calculation will be performed.

25

Referring now to Figure 4B, this figure illustrates a method for calculating an optimum display size based upon a visual object transmission rate. As noted above, certain steps can be performed in a different order with out departing from the spirit

30

and scope of the present invention. The process depicted in Figure 4B differs slightly from the process described in Figure 4A.

Only the differences between Figures 4A and 4B will be discussed with respect to Figure 4B. Accordingly, in Figure 4B, steps 410A through 435 have been
5 deleted and replaced with step 437 where the optimum display size is calculated only as a function of the visual object transmission rate.

Referring now to Figure 4C, this figure illustrates a method for calculating an optimum display size based upon a coding difficulty value alone. As noted above, certain steps can be performed in a different order with out departing from the spirit
10 and scope of the present invention. The process depicted in Figure 4C differs slightly from the process described in Figure 4A.

Only the differences between Figures 4A and 4C will be discussed with respect to Figure 4C. Accordingly, in Figure 4C, step 435 has been replaced with step
15 438 where the optimum display size is calculated only as a function of the coding difficulty value.

It is noted that if the coding difficulty calculation is performed prior to step 405A of Figure 4, then steps 405A through 430A could be eliminated since the coding difficulty value could be received with the encoded visual object.

It should be understood that the foregoing relates only to illustrative
20 embodiments of the present invention, and that numerous changes may be made therein without departing from the spirit and scope of the invention as defined by the following claims.